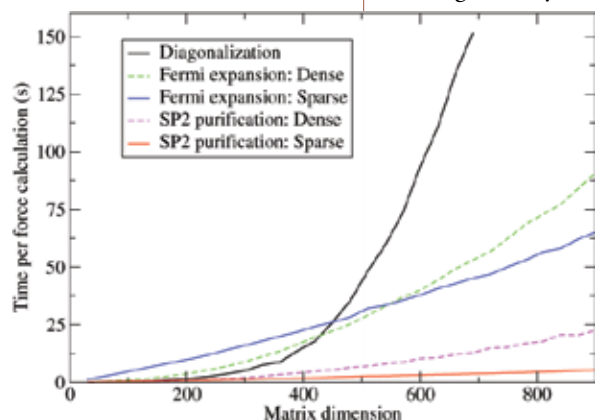


Linear Scaling Quantum Molecular Dynamics for Organic Molecules

Edward Sanville, Nicolas Bock, Matt Challacombe, Anders M. N. Niklasson, Marc Cawkwell, T-1

**For more information contact
Marc Cawkwell at
cawkwell@lanl.gov.**

Fig. 1. CPU time per MD time step as a function of matrix dimension using diagonalization of the Hamiltonian matrix, dense and sparse matrix implementations of SP2 purification, and the recursive expansion of the Fermi operator to calculate the density matrix.



Interatomic bonding in organic molecular crystals involves short-range, angularly dependent, intramolecular covalent bonds, medium-range van der Waals bonding, and long-range electrostatic interactions that arise from atomic partial charges. A predictive model for the structure and chemistry (i.e., the making and breaking of covalent bonds) of molecular crystals must describe all of these interactions, and their interdependencies, with physical rigor while ensuring computational tractability.

The selection of a mathematical model for interatomic bonding in materials typically involves a compromise between accuracy and computational speed. For example, ab initio quantum mechanical methods based on the self-consistent field or density functional theories are extremely accurate but computationally very expensive. Furthermore, the computational cost of these methods usually scales with the cube of the number of atoms, $O(N^3)$, making large-scale, long duration molecular dynamics (MD) simulations impossible. Parallel implementations of empirical interatomic potentials facilitate long duration MD simulations with many millions of atoms [1], but the ability of empirical potentials to describe a given material with high fidelity is often questionable. Furthermore, empirical

potentials tend to exhibit poor transferability to structures that differ from those employed in their parameterization.

In response to the limitations inherent to these two extremes, we have developed an $O(N)$ formalism and a supporting code that aims to strike a compromise between computational speed and physical accuracy. While at first it may seem strange to wish to compromise on the physical accuracy of an interatomic potential, a formalism that

is limited severely in the number of atoms that can be employed in a simulation will inevitably lead to errors associated with finite size effects. In many situations, errors originating from finite size effects can be more severe than those originating from approximations made during the construction of the interatomic potential.

We have generated a set of interatomic potentials for hydrocarbons based on the self-consistent charge transfer tight-binding (SCC-TB) approximation to density functional theory [2,3]. The SCC-TB formalism captures naturally the formation of covalent bonds and the transfer of charge between species of different electronegativity. The SCC-TB method is explicitly quantum mechanical since we construct and solve a time-independent Schrödinger equation for a system of interacting, atom-centered valence electrons, but the elements of the Hamiltonian matrix are parameterized rather than computed exactly. Semi-empirical potentials such as those based on tight-binding tend to show better transferability than empirical potentials owing to their more rigorous theoretical underpinning. They are also orders of magnitude faster than ab initio methods since many difficult-to-calculate terms are approximated and parameterized [4].

Our SCC-TB formalism is expressed in a minimal basis of an orthogonal set of real free-atom-like orbitals. The non-spin polarized SCC-TB Hamiltonian, H , is a sum of a Slater-Koster Hamiltonian that describes the formation of covalent bonds, H^0 , and the electrostatic potential generated by partial charges, q_i , on each atom,

$$H_{i\alpha,j\beta} = H_{i\alpha,j\beta}^0 + \frac{1}{2} \sum_k \gamma_{ik} q_k \delta_{ij} \delta_{\alpha\beta}$$

where i and j label atoms, α and β label orbitals, s, p_x, p_y, p_z , etc., γ_{ii} is the Hubbard U for species i , and $\gamma_{ij} = 1/R_{ij}$, where R_{ij} is the scalar distance between atoms i and j . The total energy is then

$$E_{\text{total}} = 2\text{tr}[\rho H] - \frac{1}{2} \sum_{i,j} \gamma_{ij} q_i q_j + E_{\text{pair}}$$

where ρ is the density matrix calculated from H and the electronic occupation. The SCC-TB equations must be solved self-consistently since the partial charges on which the Hamiltonian depends are obtained from the density matrix, i.e., $q_i = 2 \sum_{\alpha \in i} [\rho_{i\alpha,i\alpha} - \rho_{i\alpha,i\alpha}^0]$, where ρ^0 is the density matrix for isolated atoms. The last term in the

expression for the total energy, E_{pair} , is a sum of pair potentials that are used to provide short-range repulsion and $-1/R^6$ van der Waals interactions at long range.

In contrast to ab initio electronic structure methods, the construction of the Hamiltonian in SCC-TB is computationally very cheap. Instead, the bottleneck in SCC-TB calculations is the $O(N^3)$ computation of the density matrix. In the Los Alamos Transferable Tight-binding for Energetics (LATTE) code, we have implemented a number of methods for rapid, $O(N)$ calculations of the density matrix [5].

The most straightforward method for calculating the density matrix involves the diagonalization of the Hamiltonian [6]. While this method is simple to implement and enables a finite electron temperature to be easily incorporated, it is unsuited to calculations including large numbers of atoms because of $O(N^3)$ scaling (see Fig. 1). We find that Niklasson's second-order spectral projection (SP2) density matrix purification method is by far the most efficient method for the computation of the density matrix for large systems [7]. The starting point for SP2 purification is a remapping of all eigenvalues of the Hamiltonian into the interval $[0,1]$. This matrix is purified such that the eigenvalues of occupied states are projected toward 1 and those of unoccupied states toward zero until an idempotent density matrix is obtained. The purification operations require only one $O(N^3)$ matrix-matrix multiplication per iteration. However, for molecular systems we take advantage of the linear increase in the number of non-zero matrix elements with the number of molecules to achieve $O(N)$ matrix-matrix multiplications. Figure 1 illustrates clearly the tremendous speed-up that density matrix purification offers with respect to diagonalization for large systems even in dense matrix form, while a sparse matrix implementation yields $O(N)$ performance.

The density matrix obtained via purification corresponds to zero electronic temperature. This may lead to numerical instabilities during a MD simulation if eigenvalues cross the chemical potential. It is usually desirable to smooth these electronic transitions through the introduction of a finite electronic temperature. We find that the

recursive expansion of the Fermi operator method described in [8] allows for the calculation of a finite temperature density matrix with $O(N)$ cost for sparse molecular systems. It is clear from Fig. 1 that the recursive expansion of the Fermi operator is somewhere between diagonalization and purification in terms of speed, and that linear scaling is readily achievable.

Like SP2 purification, the Fermi operator expansion algorithm depends only on $O(N^3)$ dense matrix-matrix multiplications. This leaves LATTE with a very well defined computational bottleneck upon which we can concentrate our efforts. Thus, a key factor in the performance of LATTE is the question of how fast one can multiply square matrices. In addition to exploiting the sparsity of density matrices for molecular materials, we have found that the utilization of hybrid computational architectures is a promising route toward maximizing the speed of matrix-matrix multiplications. In Fig. 2 we show computational time as a function of matrix dimension for the double-precision general matrix multiply (DGEMM) algorithm on the Opteron central processing units (CPU) and IBM Cell Broadband Engine (Cell BE) chips on LANL's Roadrunner supercomputer. It is evident that the use of special computational architectures for time-consuming computational operations offers very significant performance gains. In parallel to this work on Roadrunner, we are actively pursuing the use of commodity graphics cards as computational accelerators such that huge performance boosts can be obtained even on off-the-shelf desktop workstations. Of particular note in this regard is the stability of LATTE's algorithms when executed in single-precision arithmetic.

A judicious selection of a physical model for interatomic

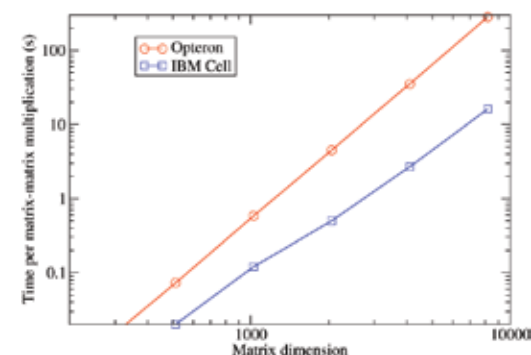


Fig 2. Time to compute a dense matrix-matrix multiplication using the DGEMM algorithm as a function of matrix dimension on the Opteron and Cell BE chips on Roadrunner.

bonding in molecular materials coupled to advanced algorithms, developments in applied mathematics, and novel computational architectures enabled us to design and run very large-scale quantum MD simulations. These simulations are currently being applied to the study of the chemistry under extreme conditions.

- [1] M.J. Cawkwell et al., *Phys. Rev. B* **78**, 014107 (2008).
- [2] M. Elstner et al., *Phys. Rev. B* **58**, 7260 (1998).
- [3] M.W. Finnis, *Interatomic Forces in Condensed Matter*, Oxford University Press (2003).
- [4] M. J. Cawkwell et al., *Phys. Rev. B* **73**, 064104 (2006).
- [5] E.J. Sanville et al., LATTE, LA-CC 10-004, Los Alamos National Laboratory.
- [6] A.P. Sutton et al., *Philos. Mag. A* **81**, 1833 (2001).
- [7] A.M.N. Niklasson *Phys. Rev. Lett. B*, **66**, 155115(2002).
- [8] A.M.N. Niklasson, *J. Chem. Phys.* **129**, 244107 (2008).

Funding Acknowledgments

- LANL Directed Research and Development Program
- DOE, NNSA Advanced Simulation and Computing (ASC) Program